PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF RECEIPT OF RECORD COPY

(PCT Rule 24.2(a))

To:

KOIKE, Akira
No. 11 Mori Building
6-4, Toranomon 2-chome
Minato-ku
Tokyo 105-0001
JAPON

Date of mailing (day/month/year)	**IMPORTANT NOTIFICATION**
21 September 2000 (21.09.00)	
Applicant's or agent's file reference	**International application No.**
SK00PCT80	PCT/JP00/05771

The applicant is hereby notified that the International Bureau has received the record copy of the international application as detailed below.

Name(s) of the applicant(s) and State(s) for which they are applicants:

SONY CORPORATION (for all designated States except US)
MIURA, Masayoshi et al (for US)

International filing date	:	25 August 2000 (25.08.00)
Priority date(s) claimed	:	26 August 1999 (26.08.99)
Date of receipt of the record copy by the International Bureau	:	12 September 2000 (12.09.00)
List of designated Offices	:	

EP :AT,BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE
National :CN,KR,US

ATTENTION

The applicant should carefully check the data appearing in this Notification. In case of any discrepancy between these data and the indications in the international application, the applicant should immediately inform the International Bureau.

In addition, the applicant's attention is drawn to the information contained in the Annex, relating to:

- [X] time limits for entry into the national phase
- [X] confirmation of precautionary designations
- [X] requirements regarding priority documents

A copy of this Notification is being sent to the receiving Office and to the International Searching Authority.

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No. (41-22) 740.14.35

Authorized officer:

Masashi HONDA

Telephone No. (41-22) 338.83.38

INFORMATION ON TIME LIMITS FOR ENTERING THE NATIONAL PHASE

The applicant is reminded that the "national phase" must be entered before each of the designated Offices indicated in the Notification of Receipt of Record Copy (Form PCT/IB/301) by paying national fees and furnishing translations, as prescribed by the applicable national laws.

The time limit for performing these procedural acts is **20 MONTHS** from the priority date or, for those designated States which the applicant elects in a demand for international preliminary examination or in a later election, **30 MONTHS** from the priority date, provided that the election is made before the expiration of 19 months from the priority date. Some designated (or elected) Offices have fixed time limits which expire even later than 20 or 30 months from the priority date. In other Offices an extension of time or grace period, in some cases upon payment of an additional fee, is available.

In addition to these procedural acts, the applicant may also have to comply with other special requirements applicable in certain Offices. **It is the applicant's responsibility** to ensure that the necessary steps to enter the national phase are taken in a timely fashion. Most designated Offices do not issue reminders to applicants in connection with the entry into the national phase.

For detailed information about the procedural acts to be performed to enter the national phase before each designated Office, the applicable time limits and possible extensions of time or grace periods, and any other requirements, see the relevant Chapters of Volume II of the PCT Applicant's Guide. Information about the requirements for filing a demand for international preliminary examination is set out in Chapter IX of Volume I of the PCT Applicant's Guide.

GR and ES became bound by PCT Chapter II on 7 September 1996 and 6 September 1997, respectively, and may, therefore, be elected in a demand or a later election filed on or after 7 September 1996 and 6 September 1997, respectively, regardless of the filing date of the international application. (See second paragraph above.)

Note that only an applicant who is a national or resident of a PCT Contracting State which is bound by Chapter II has the right to file a demand for international preliminary examination.

CONFIRMATION OF PRECAUTIONARY DESIGNATIONS

This notification lists only specific designations made under Rule 4.9(a) in the request. It is important to check that these designations are correct. Errors in designations can be corrected where precautionary designations have been made under Rule 4.9(b). The applicant is hereby reminded that any precautionary designations may be confirmed according to Rule 4.9(c) before the expiration of 15 months from the priority date. If it is not confirmed, it will automatically be regarded as withdrawn by the applicant. There will be no reminder and no invitation. Confirmation of a designation consists of the filing of a notice specifying the designated State concerned (with an indication of the kind of protection or treatment desired) and the payment of the designation and confirmation fees. Confirmation must reach the receiving Office within the 15-month time limit.

REQUIREMENTS REGARDING PRIORITY DOCUMENTS

For applicants who have not yet complied with the requirements regarding priority documents, the following is recalled.

Where the priority of an earlier national, regional or international application is claimed, the applicant must submit a copy of the said earlier application, certified by the authority with which it was filed ("the priority document") to the receiving Office (which will transmit it to the International Bureau) or directly to the International Bureau, before the expiration of 16 months from the priority date, provided that any such priority document may still be submitted to the International Bureau before that date of international publication of the international application, in which case that document will be considered to have been received by the International Bureau on the last day of the 16-month time limit (Rule 17.1(a)).

Where the priority document is issued by the receiving Office, the applicant may, instead of submitting the priority document, request the receiving Office to prepare and transmit the priority document to the International Bureau. Such request must be made before the expiration of the 16-month time limit and may be subjected by the receiving Office to the payment of a fee (Rule 17.1(b)).

If the priority document concerned is not submitted to the International Bureau or if the request to the receiving Office to prepare and transmit the priority document has not been made (and the corresponding fee, if any, paid) within the applicable time limit indicated under the preceding paragraphs, any designated State may disregard the priority claim, provided that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity to furnish the priority document within a time limit which is reasonable under the circumstances.

Where several priorities are claimed, the priority date to be considered for the purposes of computing the 16-month time limit is the filing date of the earliest application whose priority is claimed.

PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION CONCERNING SUBMISSION OR TRANSMITTAL OF PRIORITY DOCUMENT

(PCT Administrative Instructions, Section 411)

To:

KOIKE, Akira
No. 11 Mori Building
6-4, Toranomon 2-chome
Minato-ku
Tokyo 105-0001
JAPON

Date of mailing (day/month/year) 03 November 2000 (03.11.00)	
Applicant's or agent's file reference SK00PCT80	**IMPORTANT NOTIFICATION**
International application No. PCT/JP00/05771	International filing date (day/month/year) 25 August 2000 (25.08.00)
International publication date (day/month/year) Not yet published	Priority date (day/month/year) 26 August 1999 (26.08.99)
Applicant SONY CORPORATION et al	

1. The applicant is hereby notified of the date of receipt (except where the letters "NR" appear in the right-hand column) by the International Bureau of the priority document(s) relating to the earlier application(s) indicated below. Unless otherwise indicated by an asterisk appearing next to a date of receipt, or by the letters "NR", in the right-hand column, the priority document concerned was submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b).

2. This updates and replaces any previously issued notification concerning submission or transmittal of priority documents.

3. An **asterisk(*)** appearing next to a date of receipt, in the right-hand column, denotes a priority document submitted or transmitted to the International Bureau but not in compliance with Rule 17.1(a) or (b). In such a case, **the attention of the applicant is directed** to Rule 17.1(c) which provides that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity, upon entry into the national phase, to furnish the priority document within a time limit which is reasonable under the circumstances.

4. The **letters "NR"** appearing in the right-hand column denote a priority document which was not received by the International Bureau or which the applicant did not request the receiving Office to prepare and transmit to the International Bureau, as provided by Rule 17.1(a) or (b), respectively. In such a case, **the attention of the applicant is directed** to Rule 17.1(c) which provides that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity, upon entry into the national phase, to furnish the priority document within a time limit which is reasonable under the circumstances.

Priority date	Priority application No.	Country or regional Office or PCT receiving Office	Date of receipt of priority document
26 Augu 1999 (26.08.99)	11/239145	JP	13 Octo 2000 (13.10.00)

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No. (41-22) 740.14.35

Authorized officer

Magda BOUACHA

Telephone No. (41-22) 338.83.38

WO 01/16935
PCT/JP00/057

PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

To:
KOIKE, Akira
No. 11 Mori Building
6-4, Toranomon 2-chome
Minato-ku
Tokyo 105-0001
JAPON

NOTICE INFORMING THE APPLICANT OF THE COMMUNICATION OF THE INTERNATIONAL APPLICATION TO THE DESIGNATED OFFICES

(PCT Rule 47.1(c), first sentence)

Date of mailing (day/month/year) 08 March 2001 (08.03.01)	
Applicant's or agent's file reference SK00PCT80	**IMPORTANT NOTICE**
International application No. PCT/JP00/05771	**International filing date** (day/month/year) 25 August 2000 (25.08.00)
Priority date (day/month/year) 26 August 1999 (26.08.99)	
Applicant SONY CORPORATION et al	

1. Notice is hereby given that the International Bureau has communicated, as provided in Article 20, the international application to the following designated Offices on the date indicated above as the date of mailing of this Notice:
   KR,US

   In accordance with Rule 47.1(c), third sentence, those Offices will accept the present Notice as conclusive evidence that the communication of the international application has duly taken place on the date of mailing indicated above and no copy of the international application is required to be furnished by the applicant to the designated Office(s).

2. The following designated Offices have waived the requirement for such a communication at this time:
   CN,EP

   The communication will be made to those Offices only upon their request. Furthermore, those Offices do not require the applicant to furnish a copy of the international application (Rule 49.1(a-bis)).

3. Enclosed with this Notice is a copy of the international application as published by the International Bureau on
   08 March 2001 (08.03.01) under No. WO 01/16935

REMINDER REGARDING CHAPTER II (Article 31(2)(a) and Rule 54.2)

If the applicant wishes to postpone entry into the national phase until 30 months (or later in some Offices) from the priority date, **a demand for international preliminary examination** must be filed with the competent International Preliminary Examining Authority before the expiration of 19 months from the priority date.

It is the applicant's sole responsibility to monitor the 19-month time limit.

Note that only an applicant who is a national or resident of a PCT Contracting State which is bound by Chapter II has the right to file a demand for international preliminary examination.

REMINDER REGARDING ENTRY INTO THE NATIONAL PHASE (Article 22 or 39(1))

If the applicant wishes to proceed with the international application in the **national phase**, he must, within 20 months or 30 months, or later in some Offices, perform the acts referred to therein before each designated or elected Office.

For further important information on the time limits and acts to be performed for entering the national phase, see the Annex to Form PCT/IB/301 (Notification of Receipt of Record Copy) and Volume II of the PCT Applicant's Guide.

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No. (41-22) 740.14.35

Authorized officer
J. Zahra

Telephone No. (41-22) 338.83.38

09/830222

特許協力条約に基づく国際出願願書 SK00PCT80

副本 - 印刷日時 2000年08月25日 (25.08.2000) 金曜日 15時54分12秒

0	受理官庁記入欄	
0-1	国際出願番号.	
0-2	国際出願日	PCT 25.8.00 受領印
0-3	(受付印)	
0-4	様式-PCT/RO/101 この特許協力条約に基づく国際出願願書は、	
0-4-1	右記によって作成された。	PCT-EASY Version 2.91 (updated 01.07.2000)
0-5	申立て 出願人は、この国際出願が特許協力条約に従って処理されることを請求する。	
0-6	出願人によって指定された受理官庁	日本国特許庁 (RO/JP)
0-7	出願人又は代理人の書類記号	SK00PCT80
I	発明の名称	情報の検索処理方法、検索処理装置、蓄積方法及び蓄積装置
II	出願人	
II-1	この欄に記載した者は	出願人である (applicant only)
II-2	右の指定国についての出願人である。	米国を除くすべての指定国 (all designated States except US)
II-4ja	名称	ソニー株式会社
II-4en	Name	SONY CORPORATION
II-5ja	あて名:	141-0001 日本国 東京都 品川区 北品川６丁目７番３５号
II-5en	Address:	7-35, Kitashinagawa 6-chome Shinagawa-ku, Tokyo 141-0001 Japan
II-6	国籍 (国名)	日本国 JP
II-7	住所 (国名)	日本国 JP
III-1	その他の出願人又は発明者	
III-1-1	この欄に記載した者は	出願人及び発明者である (applicant and inventor)
III-1-2	右の指定国についての出願人である。	米国のみ (US only)
III-1-4ja	氏名(姓名)	三浦 雅美
III-1-4en	Name (LAST, First)	MIURA, Masayoshi
III-1-5ja	あて名:	141-0001 日本国 東京都 品川区 北品川６丁目７番３５号 ソニー株式会社内
III-1-5en	Address:	c/o SONY CORPORATION 7-35, Kitashinagawa 6-chome Shinagawa-ku, Tokyo 141-0001 Japan
III-1-6	国籍 (国名)	日本国 JP
III-1-7	住所 (国名)	日本国 JP

特許協力条約に基づく国際出願願書 SK00PCT80
副本 - 印刷日時 2000年08月25日 (25.08.2000) 金曜日 15時54分12秒

III-2	その他の出願人又は発明者	
III-2-1	この欄に記載した者は	出願人及び発明者である (applicant and inventor)
III-2-2	右の指定国についての出願人である。	米国のみ (US only)
III-2-4ja	氏名(姓名)	矢部 進
III-2-4en	Name (LAST, First)	YABE, Susumu
III-2-5ja	あて名:	141-0001 日本国 東京都 品川区 北品川６丁目７番３５号 ソニー株式会社内
III-2-5en	Address:	c/o SONY CORPORATION 7-35, Kitashinagawa 6-chome Shinagawa-ku, Tokyo 141-0001 Japan
III-2-6	国籍 (国名)	日本国 JP
III-2-7	住所 (国名)	日本国 JP
IV-1	代理人又は共通の代表者、通知のあて名 下記の者は国際機関において右記のごとく出願人のために行動する。	代理人 (agent)
IV-1-1ja	氏名(姓名)	小池 晃
IV-1-1en	Name (LAST, First)	KOIKE, Akira
IV-1-2ja	あて名:	105-0001 日本国 東京都 港区 虎ノ門二丁目６番４号 第１１森ビル
IV-1-2en	Address:	No.11 Mori Bldg., 6-4, Toranomon 2-chome Minato-ku, Tokyo 105-0001 Japan
IV-1-3	電話番号	03-3508-8266
IV-1-4	ファクシミリ番号	03-3508-0439
IV-2	その他の代理人	筆頭代理人と同じあて名を有する代理人 (additional agent(s) with same address as first named agent)
IV-2-1ja	氏名	田村 榮一; 伊賀 誠司
IV-2-1en	Name(s)	TAMURA, Eiichi; IGA, Seiji
V	国の指定	
V-1	広域特許 (他の種類の保護又は取扱いを求める場合には括弧内に記載する。)	EP: AT BE CH&LI CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE 及びヨーロッパ特許条約と特許協力条約の締約国である他の国
V-2	国内特許 (他の種類の保護又は取扱いを求める場合には括弧内に記載する。)	CN KR US

特許協力条約に基づく国際出願願書 SK00PCT80
副本 - 印刷日時 2000年08月25日 (25.08.2000) 金曜日 15時54分12秒

V-5	指定の確認の宣言 出願人は、上記の指定に加えて、規則4.9(b)の規定に基づき、特許協力条約のもとで認められる他の全ての国の指定を行う。ただし、V-6欄に示した国の指定を除く。出願人は、これらの追加される指定が確認を条件としていること、並びに優先日から15月が経過する前にその確認がなされない指定は、この期間の経過時に、出願人によって取り下げられたものとみなされることを宣言する。		
V-6	指定の確認から除かれる国	なし (NONE)	
VI-1	先の国内出願に基づく優先権主張		
VI-1-1	先の出願日	1999年08月26日 (26.08.1999)	
VI-1-2	先の出願番号	平成１１年特許願第２３９１４５号	
VI-1-3	国名	日本国 JP	
VI-2	優先権証明書送付の請求 上記の先の出願のうち、右記の番号のものについては、出願書類の認証謄本を作成し国際事務局へ送付することを、受理官庁に対して請求している。	VI-1	
VII-1	特定された国際調査機関(ISA)	日本国特許庁 (ISA/JP)	
VIII	照合欄	用紙の枚数	添付された電子データ
VIII-1	願書	4	-
VIII-2	明細書	55	-
VIII-3	請求の範囲	13	-
VIII-4	要約	1	absk00pct80.txt
VIII-5	図面	15	-
VIII-7	合計	88	
	添付書類	添付	添付された電子データ
VIII-8	手数料計算用紙	✓	-
VIII-10	包括委任状の写し	✓	-
VIII-16	PCT-EASYディスク	-	フレキシブルディスク
VIII-17	その他	納付する手数料に相当する特許印紙を貼付した書面	-
VIII-17	その他	国際事務局の口座への振込を証明する書面	-
VIII-18	要約書とともに提示する図の番号	1	
VIII-19	国際出願の使用言語名:	日本語 (Japanese)	
IX-1	提出者の記名押印		
IX-1-1	氏名(姓名)	小池 晃	

特許協力条約に基づく国際出願願書 SK00PCT80

副本 - 印刷日時 2000年08月25日 (25.08.2000) 金曜日 15時54分12秒

IX-2	提出者の記名押印	
IX-2-1	氏名(姓名)	田村 榮一
IX-3	提出者の記名押印	
IX-3-1	氏名(姓名)	伊賀 誠司

受理官庁記入欄

10-1	国際出願として提出された書類の実際の受理の日	
10-2 10-2-1 10-2-2	図面 : 受理された 不足図面がある	
10-3	国際出願として提出された書類を補完する書類又は図面であってその後期間内に提出されたものの実際の受理の日 (訂正日)	
10-4	特許協力条約第11条(2)に基づく必要な補完の期間内の受理の日	
10-5	出願人により特定された国際調査機関	ISA/JP
10-6	調査手数料未払いにつき、国際調査機関に調査用写しを送付していない	

国際事務局記入欄

11-1	記録原本の受理の日	

EP ・ US

PCT

特 許 協 力 条 約

国際調査報告

（法８条、法施行規則第40、41条）
〔ＰＣＴ１８条、ＰＣＴ規則43、44〕

出願人又は代理人の書類記号 SK00PCT80	今後の手続きについては、国際調査報告の送付通知様式（ＰＣＴ／ＩＳＡ／２２０）及び下記５を参照すること。	
国際出願番号 ＰＣＴ／ＪＰ００／０５７７１	国際出願日 (日.月.年) ２５．０８．００	優先日 (日.月.年) ２６．０８．９９
出願人（氏名又は名称） ソニー株式会社		

国際調査機関が作成したこの国際調査報告を法施行規則第41条（ＰＣＴ１８条）の規定に従い出願人に送付する。
この写しは国際事務局にも送付される。

この国際調査報告は、全部で 3 ページである。

☐ この調査報告に引用された先行技術文献の写しも添付されている。

１．国際調査報告の基礎

ａ．言語は、下記に示す場合を除くほか、この国際出願がされたものに基づき国際調査を行った。
☐ この国際調査機関に提出された国際出願の翻訳文に基づき国際調査を行った。

ｂ．この国際出願は、ヌクレオチド又はアミノ酸配列を含んでおり、次の配列表に基づき国際調査を行った。
☐ この国際出願に含まれる書面による配列表
☐ この国際出願と共に提出されたフレキシブルディスクによる配列表
☐ 出願後に、この国際調査機関に提出された書面による配列表
☐ 出願後に、この国際調査機関に提出されたフレキシブルディスクによる配列表
☐ 出願後に提出した書面による配列表が出願時における国際出願の開示の範囲を超える事項を含まない旨の陳述書の提出があった。
☐ 書面による配列表に記載した配列とフレキシブルディスクによる配列表に記録した配列が同一である旨の陳述書の提出があった。

２．☐ 請求の範囲の一部の調査ができない（第Ⅰ欄参照）。

３．☐ 発明の単一性が欠如している（第Ⅱ欄参照）。

４．発明の名称は
☒ 出願人が提出したものを承認する。
☐ 次に示すように国際調査機関が作成した。

５．要約は
☒ 出願人が提出したものを承認する。
☐ 第Ⅲ欄に示されているように、法施行規則第47条（ＰＣＴ規則38.2(b)）の規定により国際調査機関が作成した。出願人は、この国際調査報告の発送の日から１カ月以内にこの国際調査機関に意見を提出することができる。

６．要約書とともに公表される図は、
第 1 図とする。☒ 出願人が示したとおりである。 ☐ なし
☐ 出願人は図を示さなかった。
☐ 本図は発明の特徴を一層よく表している。

国際調査[redacted]　　国際出願番[redacted]PCT／JP00／05771

A.　発明の属する分野の分類（国際特許分類（IPC））

Int Cl' G10L15／00,　G11B27／30,　H04N5／76, H04N5／91

B.　調査を行った分野

調査を行った最小限資料（国際特許分類（IPC））

Int Cl' G10L15／00～17／00,　G11B27／10～27／32;
H04N5／76, H04N5／91

最小限資料以外の資料で調査を行った分野に含まれるもの
日本国実用新案公報　　1926～1995年
日本国公開実用新案公報　1971～2000年
日本国登録実用新案公報　1994～2000年
日本国実用新案登録公報　1996～2000年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）
JICST科学技術文献ファイル（JOIS）　　INSPEC（DIALOG）
WPI（DIALOG）
IEEE／IEE Electronic Library

C.　関連すると認められる文献

引用文献のカテゴリー＊	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求の範囲の番号
A	Proceedings of 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, Vol.6, T.Lambrou et al, "Classification of audio signals using statistical features on time and wavelet transform domains", p.3621-3624, 12-15 May 1998, ISBN0-7803-4428-6, IEEE Catalog Number 98CH36181	1～47
A	Proceedings of 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, Vol.2, H.Soltau et al, "Recognition of music types", p.1137-1140, 12-15 May 1998, ISBN 0-7803-4428-6, IEEE Catalog Number 98CH36181	1～47

☒ C欄の続きにも文献が列挙されている。　　☐ パテントファミリーに関する別紙を参照。

＊　引用文献のカテゴリー
「A」特に関連のある文献ではなく、一般的技術水準を示すもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献
「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」同一パテントファミリー文献

国際調査を完了した日	国際調査報告の発送日
11.10.00	07.11.00

国際調査機関の名称及びあて先	特許庁審査官（権限のある職員）	
日本国特許庁（ISA／JP） 郵便番号100－8915 東京都千代田区霞が関三丁目4番3号	松尾　淳一　印	5C　8842
	電話番号　03－3581－1101　内線　3540	

C（続き）．　関連すると認められる文献

引用文献のカテゴリー＊	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求の範囲の番号
A	Proceedings of the Fifth International Symposium on Signal Processing and Its Applications, ISSPA'99, Vol.1, K.Melih et al, "Audio source type segmentation using a perceptually based representation", p.51-54, 22-25 Aug. 1999, ISBN1-86435-451-8, IEEE Catalog Number 99EX359	1～47
A	電子情報通信学会論文誌(Transactions of the Institute of Electronics, Information and Communication Engineers), Vol.J79-D-II, No.11, November 1996, 柏野邦夫外(Kunio Kashiwano et al), 「音楽情景分析の処理モデルOPTIMAにおける単音の認識」("Note recognition mechanisms in the OPTIMA processing architecture for music scene analysis"), p.1751-1761, 1996年11月25日発行(25.11.96), ISSN0915-1923	1～47
A	情報処理学会研究報告[音楽情報科学](IPSJ Sig Notes[SIGMUS]), Vol.97, No.18, 97-MUS-19-11, 柏野邦夫外(Kunio Kashiwano et al), 「適応型混合テンプレートを用いた音源同定-複数楽器演奏への適用-」("Sound source identification using adaptive template mixtures - Formulation and application to music stream segregation -"), p.55-68, 1997年2月20日発行(20.02.97), ISSN0919-6072	1～47
A	ＪＰ，９－９１９９，Ａ(ソニー株式会社) 10.1月.1999(10.01.99) 全文全図（ファミリーなし）	1～47
A	ＪＰ，５－３３４８６１，Ａ(日本無線株式会社) 17.12月.1993 (17.12.93) 全文全図（ファミリーなし）	1～47
A	ＪＰ，７－１０５２３５，Ａ(シャープ株式会社) 21.4月.1995 (21.04.95) 全文全図 &JP,3021252,B2	1～47
A	ＪＰ，８－２６５６６０，Ａ(日本電信電話株式会社) 10.11月.1996(10.11.96) 全文全図（ファミリーなし）	1～47
A	ＪＰ，１０－３０７５８０，Ａ(日本電信電話株式会社) 17.11月.1998(17.11.98) 全文全図（ファミリーなし）	1～47
A	ＪＰ，１０－３１９９４８，Ａ(日本電信電話株式会社) 4.12月.1998(04.12.98) 全文全図（ファミリーなし）	1～47

論　文

音楽情景分析の処理モデルOPTIMAにおける単音の認識

柏野　邦夫[†*]　　中臺　一博[†**]　　木下　智義[†]　　田中　英彦[†]

Note Recognition Mechanisms in the OPTIMA Processing Architecture for Music Scene Analysis

Kunio KASHINO[†*], Kazuhiro NAKADAI[†**], Tomoyoshi KINOSHITA[†], and Hidehiko TANAKA[†]

あらまし　音楽演奏の音響信号を対象として演奏情報を認識する試みとしては，従来自動採譜の研究が行われているが，複数種類の楽器音を含む音楽演奏を対象とする場合には，認識処理の有効性は極めて限られていた．そこで本論文では，複数種類の楽器音を含む音楽演奏の認識を音楽情景分析の問題としてとらえ，その解決を図る．ここで音楽情景分析とは，音楽演奏の音響信号から，単音や和音などの音楽演奏情報を記号表現として抽出することを指す．本論文ではまず，音楽情景分析を実現する上では情報統合の技術が不可欠であるとの認識から，ベイジアンネットワークによる情報統合の機構を備えた音楽情景分析の処理モデルOPTIMAを提案する．次に，特に単音の認識に的を絞って，提案する情報統合機構の有効性を示す．

キーワード　聴覚的情景分析，音源分離，情報統合，ベイジアンネットワーク，自動採譜

1. まえがき

本論文では，情報統合の機構を備えた音楽情景分析の処理モデルを提案し，特に単音の認識にかかわる処理に着目して，情報統合の有効性を示す．

一般に情景分析とは，感覚情報を入力として外界に生じている事象や外界に存在する物体に関する記述を出力する情報処理のことをいう．従来，情景分析の研究は主に画像情報を対象としていたが，近年，さまざまな音響情報の認識を情景分析の観点からとらえる考え方，すなわち聴覚的情景分析（auditory scene analysis）の枠組みが提案された[1]．聴覚的情景分析のうちで特に音楽音響信号を対象とするものを，本論文では音楽情景分析（music scene analysis）と呼ぶ．具体的には，音楽情景分析とは，音楽音響信号を入力とし，各楽器の演奏情報（単音，和音，リズムなど）を記号表現として出力する情報処理を指す．

計算機への音楽の演奏情報の入力に関しては，これまでにも自動採譜システムの研究が行われている[2]～[6]．しかし従来は，単一楽器の単旋律（ソロ歌唱など）か，または単一楽器の多重音（ピアノ演奏など）を対象とした研究が主であった．複数楽器の多重音を対象とする研究も試みられてはきたが[7]～[9]，音源の分離と同定が問題となるために，認識処理の有効性は限られていた．

そこで本論文では，複数楽器の多重音を含む音楽演奏を対象とする演奏情報の認識の問題を，聴覚的情景分析の観点からとらえて解決を図る．聴覚的情景分析において重要なのは，入力信号だけでなく，対象に関するモデルや統計的データなど利用可能なさまざまな情報を統合して総合的な判断を行うことである[10]．よって本論文では，情報統合の機構を示すこと，およびその機構の単音の認識に関する有効性を示すことの2点を目的とする．以下，まず2.において，提案する音楽情景分析の処理モデルの全体像を示し，3.において情報統合の機構を説明する．次に，4.と5.において，特に単音の認識に着目して，ボトムアップ処理の概要とトップダウン処理の概要をそれぞれ説明する．6.でシステムの動作例を示した後，7.において，単音の認識に関する評価実験を行い，ボトムアップ処理のみによる実験結果と，ボトムアップ処理とトップダウン処理とを併用した場合の実験結果とを比較することによって，提案する処理モデルの単音認識に対する有

† 東京大学工学部電気工学科，東京都
Faculty of Engineering, University of Tokyo, Bunkyo-ku, Tokyo, 113 Japan
* 現在，NTT基礎研究所
** 現在，NTTソフトウエア本部

効性を示す．8. をむすびとする．

2. 処理モデル OPTIMA の全体像

2.1 構　　成

提案する音楽情景分析の処理モデル OPTIMA (Organized Processing toward Intelligent Music Scene Analysis) の全体像を図 1 に示す．OPTIMA は，各時点で得られた情報に基づいて，周波数成分 (frequency component)，単音 (musical note)，および和音 (chord) についての仮説を生成し，事後確率最大を評価基準として，全体として最ももっともらしい仮説の組を逐次求めていく枠組みである．図 1 に示すシステムの入力はモノラルの音楽音響信号であり，出力は，和音記号の列，楽器ごとに分類された単音記号の列，楽器ごとに分類された周波数成分の組，および拍位置を表す記号列である．これらの記号列は，音楽演奏に対する「知覚的な音」[10] に相当する．なお，出力される周波数成分をもとに，楽器ごとの音響信号波形を再合成することも可能である．

本処理モデルは，(A) 前処理部 (preprocesses)，(B) 主処理部 (main processes)，(C) 知識源 (knowledge sources)，および (D) 出力データ生成部 (output data generation) の四つの部からなる．

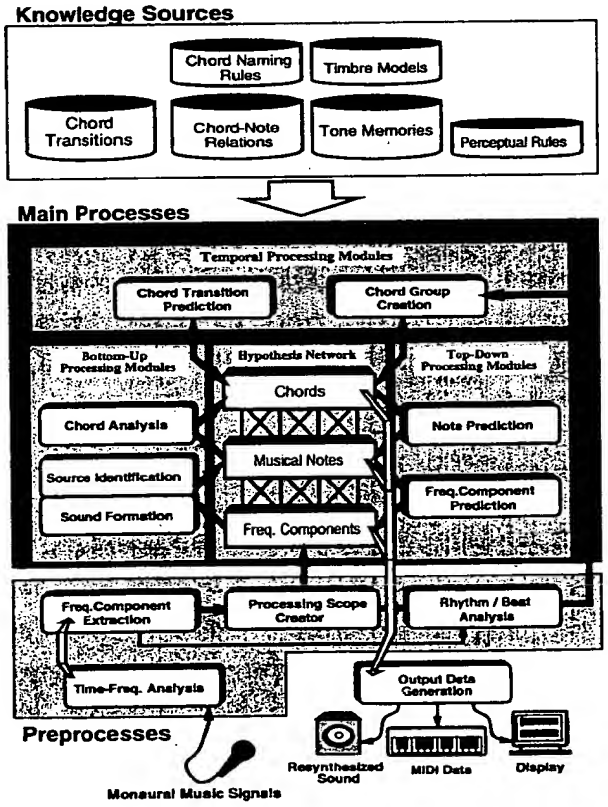


図 1　処理モデル OPTIMA の全体像
Fig. 1　The OPTIMA processing architecture.

前処理部は，入力音響信号を時間と周波数に関するエネルギー表現に変換すると共に，このエネルギー表現上における特徴を周波数成分として抽出し，リズム情報によりこれを整形して，主処理部に対する入力となる処理単位 (processing scope) を形成する部分である．ここで周波数成分とは，サウンドスペクトログラム上で，時間的に連続した，周波数方向に見たときのパワーの極大点の集合をいう．また処理単位とは，立上り時刻が互いに近接した周波数成分の集合を指す．

主処理部は，音響事象の仮説を保持するためのベイジアンネットワーク (仮説ネットワーク；hypothesis network) を備えている．仮説ネットワークは，(1) 周波数成分，(2) 単音，および (3) 和音の三つの抽象度の階層をもつ．単音は，個々の音符に対応する記号表現である．和音は，時間的に近接した複数の単音によって特徴づけられる記号表現である．仮説ネットワークに対して，(a) 抽象度の低い階層から抽象度の高い階層への情報表現の変換を行うボトムアップ処理モジュール (bottom-up processing modules)，(b) 抽象度の高い階層から抽象度の低い階層への情報表現の変換を行うトップダウン処理モジュール (top-down processing modules)，(c) 時間の推移に関する情報を扱う処理モジュール (temporal processing modules)，の三つの群に分けられる処理モジュールが情報を書き込む．ボトムアップ処理モジュールとしては，周波数成分の情報をもとに単音の情報を生成する処理 (単音仮説生成；sound formation および source identification)，単音の情報をもとに和音の情報を生成する処理 (和音仮説生成；chord recognition) の二つがある．トップダウン処理モジュールとしては，和音の情報をもとに単音仮説の確からしさに関する情報を出力する処理 (和音構成音情報付与；note prediction) と，単音の情報をもとに周波数成分仮説の確からしさに関する情報を出力する処理 (単音構成周波数成分情報付与；frequency component prediction) の二つがある．また，時間方向の処理モジュールとしては，和音の推移に関する情報を出力する処理 (和音遷移情報付与；chord transition prediction) と，時間的に連続する何個の処理単位が一つの和音を形成するかに関する情報を出力する処理 (chord group creation) の二つがある．これらの処理モジュールのうち本論文で扱うものは，単音仮説生成

と単音構成周波数成分情報付与の二つである．

主処理部における各処理モジュールは，それぞれ必要に応じて知識源を参照する．知識源としては，和音遷移に関する統計データ（和音遷移情報；chord transitions），和音を構成する単音に関する統計データ（和音構成音情報；chord-note relations），単音の集合に対しどのような和音名をつけるかをルールとしたもの（和音名ルール；chord naming rules），単音を構成する周波数成分に関するデータ（単音記憶；tone memories），音色を表現する特徴空間（音色モデル；timbre models），および単音形成のための知覚的ルール（perceptual rules）を備える．これらのうち本論文で扱うものは，単音仮説生成モジュールの参照する知覚的ルールおよび音色モデル，単音構成周波数成分情報付与モジュールの参照する単音記憶である．

出力データ生成部は，主処理部の仮説ネットワークにおいて事後確率最大となった仮説を，画面表示やMIDI（Musical Instrument Digital Interface）データなど目的に応じた形で出力するためのものである．

2.2 他の聴覚的情景分析のモデルとの比較

聴覚的情景分析に関する研究[10]のうちで，ボトムアップ処理だけではなく種々の情報の統合を考慮したものとしては，中谷らによる音響ストリーム分離の試みと，Lesserらによる IPUS（Integrated Processing and Understanding of Signals）プロジェクトが挙げられる．

中谷らの実験システムは，単純な機能をもつ複数のエージェントから構成されるマルチエージェントシステムである[13],[14]．複数の処理モジュールの協調によって処理を行うという点では，中谷らのモデルと本モデルとは共通している．但し，中谷らのモデルが，自由度が高い反面で数量的定式化の困難な情報処理アーキテクチャを背景としているのに対し，本モデルは，情報統合の方法としてベイズの定理に基づく確率モデルを用いている．動作の理解と解析の容易さという点では，本モデルの方が見通しが良い．また，中谷らのモデルの出力が音源（話者）ごとに分離した音響信号であるのに対し，本論文のモデルでは，音響信号をもとに記号表現を生成して出力する点が異なっている．

またLesserらのIPUSシステムは，黒板モデルに基づいている[15]．本モデルも，複数のモジュールが共通の空間に対して情報の読出しおよび書込みを行うことで処理が進行するという点においては，黒板モデルと同様である．しかしLesserらの黒板モデルでは，情報の統合および処理の制御は知識源として備えられた制御ルールによって行われており，情報統合のための定量的指針がない．このため，制御の安定性や処理の有効性を確保するための制御ルールの調整や保守が容易ではない．これに対し，本モデルでは，各処理モジュールは局所的な起動条件がそろったときに起動するだけであり，グローバルな制御ルールは不要である．また処理の保守としては，処理モジュールの独立性が確保されていることから，それぞれの処理モジュールにおける精度向上のみを考えればよい．

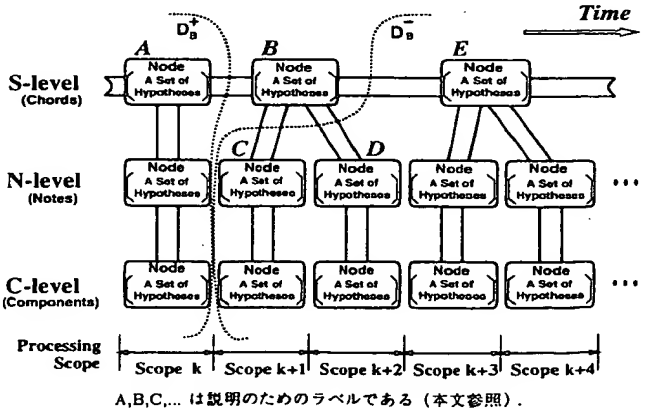
3. 情報統合のモデル

本章では，提案する処理モデルの主処理部に備わっている仮説ネットワークにおける情報統合の原理と，これに基づく処理モジュールの動作について説明する．

3.1 情報統合の原理

まず，処理の対象とする音楽演奏における抽象度の階層と時間的なつながりを考慮して，図2のような構造を考える．階層は，下から周波数成分（C）レベル，単音（N）レベル，および和音（S）レベルである．周波数成分レベルのノードと単音レベルのノードは1対1に対応するが，時間方向の複数の単音の並びが一つの和音を成し得るので，一般には単音レベルの複数のノードから和音レベルの一つのノードに対しリンクを設ける．

各ノードは，一般に複数の仮説を保持する．すなわち，周波数成分レベルのノードでは処理単位における周波数成分仮説，単音レベルのノードでは処理単位における単音仮説，和音レベルのノードでは和音のN-gramの仮説をそれぞれ保持する．これらのうち周



A,B,C,... は説明のためのラベルである（本文参照）．

図2 仮説ネットワークの構造
Fig. 2 Structure of the hypothesis network.

波数成分仮説は，処理単位における周波数成分の状態を個々の仮説としたものである．また単音仮説は，例えば「フルートの音高番号 60 とピアノの音高番号 48」といったように，各時点で同時に発音している単音の音源名と高さとを組にしたものである．

一方，各リンクは，隣接するノードにおける仮説同士の相関を，条件付確率として保持する．

以下 **3.1** の範囲の内容は，Pearl [16] に基づくものであるが，**3.2** で我々が提案する処理モジュールの構成と動作を理解する上で重要なので，要点のみを簡潔に記す．今，図 2 に示したノード B に着目する．B の子孫のノードに保持される仮説全体を D_B^- とし，B でも B の子孫でもないノードに含まれる仮説全体を D_B^+ とすれば，B に保持される仮説 $b=(b_1, b_2, \cdots, b_m)$ の確信度ベクトル $BEL(b)$ は，周囲のノードの仮説の状態が与えられた条件下での着目するノードの仮説の確率という意味で，

$$BEL(b) = P(\, b \,|D_B^+, D_B^-) \tag{1}$$

と書ける．なお本論文では，確信度という語を，式 (1) のように，各ノードにおいて保持される動的な条件付確率を指す場合に用いる．ここで

$$P(D_B^+, D_B^-|\, b) = P(D_B^+|\, b)\; P(D_B^-|\, b) \tag{2}$$

を仮定すれば（つまり，ノード B の仮説の状態が決まれば，B の親（子）側のノードの仮説の確率は B の子（親）側のノードの仮説の状態にかかわらず決まることを仮定すれば），ベイズの定理を用いて

$$P(\, b \,|D_B^+, D_B^-) = \alpha P(D_B^-|\, b)\; P(\, b \,|D_B^+) \tag{3}$$

と式変形することができる（本章においては，ベクトルの積の表記は，ベクトルの対応する要素同士の積を要素とするベクトルを得る操作を表すものとする）．ここで α は正規化定数である．$\lambda(b) = P(D_B^-|\, b)$, $\pi(b) = P(\, b \,|D_B^+)$ とおけば，

$$BEL(b) = \alpha\, \lambda(b)\, \pi(b) \tag{4}$$

となる．$\lambda(b)$ はノード B とその子側のノードとの関連を，また $\pi(b)$ はノード B とその親側のノードとの関連を表している．

式 (4) より，$BEL(b)$ を求めるためには，$\lambda(b)$ と $\pi(b)$ を求めればよい．そこでまず $\lambda(b)$ について考える．B の k 番目の子をルートとする副木に含まれる仮説を D^{k-} と書くと，

$$\lambda(b) = \beta \prod_k P(D^{k-}|\, b) \tag{5}$$

となる（β は正規化定数）．但し，親の仮説が定まったときの，子の間での仮説の独立性を仮定している．ここで，今仮に k 番目の子がノード E だったとすると，

$$P(D^{k-}|\, b) = \sum_i \lambda(e_i)\; P(e_i|\, b) \tag{6}$$

となることが示せるので，式 (5) と合わせると，親から子への条件付確率（すなわち $P(e_i|b_j)$ など）が与えられれば，漸化的に λ を伝搬できることがわかる．

次に $\pi(b)$ について考えると，

$$\pi(b) = \sum_i P(\, b \,|a_i) \left\{ \gamma \pi(a_i) \prod_m \lambda_m(a_i) \right\} \tag{7}$$

となることが示される．但し m は B を除く B の兄弟姉妹を数える添字であり，γ は正規化定数である．式 (7) の中括弧の中は，$BEL(a)$ の計算において必要なものであるから，$BEL(a)$ を計算した時点でわかっている．そこで，π についても，親から子への条件付確率（すなわち $P(b_j|a_i)$ など）が与えられれば，漸化的に π を伝搬できることがわかる．

結局，式 (4) において，親から子への条件付確率が与えられれば，確率としての性質に矛盾しない形で，双方向に確率を伝搬することによって確信度ベクトルが求まることがわかった．仮定した条件は，あるノードの仮説の状態が決まったとき，そのノードの親と子の間の独立性（式 (2)），およびそのノードの子同士の間の独立性（式 (5)）である．本論文の場合，和音レベルにおいて和音の N-gram 仮説を保持しており，また一般に単音が決まれば和音は周波数成分の状態にかかわらず定まると考えることができるので，前者の独立性の仮定は妥当なものである．一方，後者の独立性を仮定していることから，単音の時間的なつながりは考慮されていない．

音楽情景分析システムにおける情報統合の機構は，仮説生成の順序に自由度があり，処理の制御が容易であり，かつ数多くの緩い制約を効率的に扱えることが必要である．Pearl のベイジアンネットワークは，このような条件を満たしているので，音楽情景分析に適した手法である．

3.2 処理モジュールの動作

以上の原理に沿った計算を行うために，仮説ネット

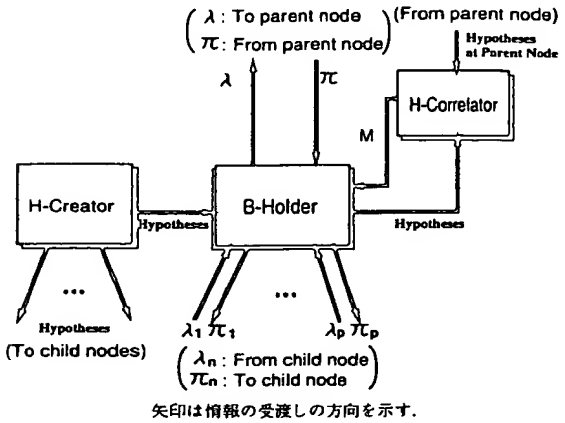


図3 1ノード当りの仮説ネットワークの構成要素
Fig.3 Modules required for a node of the hypothesis network.

ワークの各ノードにおいて次のようなモジュールを設ける．これらは，図3のように連係する．以下，着目するノードを B とし，その親を A として説明する．

（1） 確信度ホルダ（B-Holder）： 確信度ベクトル $BEL(b)$ を保持し，伝搬させる．仮説ネットワークの一つのノードに一つずつ存在する．

（2） 仮説クリエータ（H-Creator）： ある確信度ホルダに対応する仮説 b_j を作り，確信度の初期値を与える．

（3） 仮説コリレータ（H-Correlator）： 隣接する確信度ホルダに関して，$P(b_j|a_i)$ を評価する．

本論文の例では，仮説クリエータとしてボトムアップ処理モジュールを用い，仮説コリレータとしてトップダウン処理モジュールを用いている．確信度ホルダは仮説ネットワークの実体である．

このうち，確信度ホルダの動作は次のとおりである．確信度ホルダは，自分の親と子の有無と，(存在する場合には）それらとの通信のためのアドレスを認識している．親がいない確信度ホルダは，自分がルートであることを知っている．また，和音レベルの確信度ホルダは，時間方向の子と階層方向の子とを区別することができるとする．

ある処理単位における処理は，まず仮説クリエータが確信度ホルダを生成し，これに仮説を与えることにより開始する．仮説クリエータは，処理に必要なデータがそろうなど，起動可能な状態になったときに起動する．一度生成された確信度ホルダは，λ または π を受け取ることによって起動される．起動されたら，式(5)および式(7)に従って内部状態ベクトル λ または π を変更する．この結果，式(4)によって $BEL(b)$ が更新される．次に確信度ホルダは，隣接ノードの確信度ホルダに渡すべき λ と π，つまり $\lambda_B(a)$ および $\pi_k(b)$ を作成してこれを伝搬する．すなわち

$$\lambda_B(a) = M^t\lambda(b) \tag{8}$$

および

$$\pi_k(b) = \zeta\pi(b)\prod_{j\neq k}\lambda_j(b) \tag{9}$$

とする．ここで ζ は正規化定数であり，M^t は $P(b_j|a_i)$ を要素とする行列 M の転置行列を表す．M は，仮説コリレータから与えられる．すなわち，隣接するノードの確信度ホルダに対する仮説の生成が終了すると，トップダウンプロセスである仮説コリレータが起動し，知識源を参照して $P(S|S')$，$P(N|S)$，$P(C|N)$ を評価する．ここで $P(S|S')$ は和音の N-gram の遷移確率，$P(N|S)$ はある和音のときある音高の単音が出現する確率，また $P(C|N)$ はある単音のときある周波数成分の状態となる確率である．

このようにして，自分の起動の原因となったリンクは除き，他のすべての子と親に対して λ と π の伝搬が行われる．伝搬すべき相手が存在しなければ確信度ホルダは単に自らが保持している確信度ベクトルの変更のみを行い，再び隣接ノードからの λ または π によって起動されるまで休眠する．

ルートの確信度ホルダは，時間方向につながったノードの数が一定値に達した時点で時間方向の子を切り離し，自分の時間方向の子を新たなルートにすることができる．時間方向の子を切り離した確信度ホルダは，その階層方向の子孫とともに消滅する．確信度ホルダが消滅した時点で，保持されていた仮説の確信度は確定することになる．

4. 単音仮説の生成

単音レベルの仮説クリエータは，ボトムアップ処理によって単音仮説を生成する．単音仮説生成処理は，単音形成処理（単音を形成する周波数成分のクラスタリング）と，音源同定処理（各単音についての楽器種類の判別）とに分けられる[17]．以下，これらの処理について順に検討する．

4.1 単音形成処理

本論文の単音形成処理では，入力に対し次の二つを仮定して，処理単位ごとに行う．

[仮定 1] 一つの単音に含まれる任意の周波数成分は，最も低い周波数成分に対してほぼ高調波関係にあること

[仮定 2] 一つの単音に含まれるすべての周波数成分の立上り（開始端点）の時刻がほぼ同時であること

この仮定は，人間の聴覚的特性および対象とする音の性質から見て妥当なものである．すなわち音響心理学の知見によれば，人間の音源分離知覚（単音の形成）に関して，周波数成分の高調波関係の有無と周波数成分の立上り時刻の同時性の有無が音源分離知覚に強い影響を与えることがわかっている．また，音楽音響学の知見によれば，人間がピッチを知覚するような楽器音では，高調波の非調和性はおおむね3%程度以下と考えてよく，また擦弦楽器，管楽器，打弦楽器のいずれにおいても，周波数成分の立上り時刻のずれは数十ms程度以下であることがわかっている[18]．なお，一部の打楽器音などのようにピッチを有しない音については，本論文では扱わないものとする．

さて，上の二つの仮定の下での検討課題は，周波数成分に高調波関係のずれと立上り時刻のずれという複数の特徴が存在したとき，これらをいかに評価してクラスタリングを行うかである．本論文では，単音形成クラスタリングにおける評価統合モデルを用いる[17]．評価統合モデルは，まず複数の特徴が独立に評価され，次にその評価値が統合されるとするモデルであり，等振幅の二つの周波数成分だけが存在するという最も基本的な場合に関しては，聴覚実験結果と対応することが示されている[17]．文献[17]によれば，周波数成分の高調波関係のずれによって分離知覚の生じる確実性を c_h，立上り時刻のずれによって分離知覚の生じる確実性を c_o としたとき，

$$m = 1 - (1 - c_h)(1 - c_o) \tag{10}$$

によって双方の特徴が存在したときの分離知覚の確実性 m が得られるが，ここでは，m を周波数成分間の距離として単音形成のためのクラスタリングに用いる．クラスタリングは，前述の仮定1に注意すれば，次のような操作によって行うことができる．

（1） 最も低い周波数の周波数成分をクラスタ中心 C_1 とする

（2） 周波数の低い順に周波数成分を走査し，C_1 との距離が m_θ より大きい周波数成分を見出して，新たなクラスタ中心 C_2 とする

（3） いずれのクラスタ中心に対しても，距離が m_θ より大きい周波数成分を見出して，新たなクラスタ中心 C_3 とする

（4） これを新たなクラスタ中心が見出せなくなるまで繰り返す

（5） 各クラスタ中心について，距離が m_θ を超えない周波数成分すべてを見出し，それぞれのクラスタに所属させる

ここで m_θ は別の音と知覚するための確実性に対するしきい値であり，0から1までの値をとる．0に近いほどいわゆる分析的な聞き方に近づく．また5番目の操作で，立上り時刻がクラスタ中心の周波数成分の立上り時刻よりも前にある成分については，立上り時刻に関する評価値を算入しないものとする．これは重複周波数成分（shared component；複数の単音に属する周波数成分が重なったために一つの周波数成分として観測されたもの）を考慮するためである．以上のような操作により，人間が一つの音と聞く可能性の高い周波数成分がクラスタ化される．

しかし，このような操作だけでは，ある単音の基本周波数が他の単音の基本周波数の整数倍になっている場合（同一またはオクターブ差の音程）には，これらを別の単音としてクラスタ化することができない．そこで，基本周波数が整数比となるような単音が含まれる仮説をも生成するものとした．この際，無限に多くの仮説が生じないようにするため，同時に発音する単音の最大数に制限を設けた．例えば，最大同時発音数を3としたとき，音高番号60の単音に対して(60,60)，(60,60,60)，(60,72)，(60,84)，(60,72,84)，(60,60,72)，(60,72,72)，… などの単音仮説を生成する．

4.2 音源同定処理

音源同定処理は，音色空間における判別分析によって行う．音色空間は，音楽音響の分野の知見や楽器の構造等を考慮して選択した41のパラメータに関して主成分分析を行うことにより構成した．選択したパラメータの主なものを表1に示す．主成分分析の寄与率は95%とした．また，音色空間の次数を n としたとき，各音色は音色空間上で n 次元の正規分布として表すことができると仮定し，音色ごとの音色重心および分散・共分散行列を音色モデル（timbre model）として知識源に蓄積した．このとき，i 番めのサンプルが音色カテゴリー A に属する確率 P_{Ai} を式(11)で算出することができる．

$$P_{Ai} = \frac{1}{(2\pi)^{m/2}\sqrt{|S_A|}} \exp\left\{-\frac{1}{2}D_{Ai}^2\right\} \tag{11}$$

表 1　音色空間の構成に用いた特徴量の例
Table 1　Examples of parameters used for principal component analysis of timbres.

周波数成分に関する特徴量	周波数成分のパワーの比 周波数成分の立上り時刻の差 観測される周波数成分の数
パワー包絡に関する特徴量	立上りの傾き 振幅変化の度合（振幅変調度） 振幅変化の周波数

但し D^2_{Ai} はマハラノビスの汎距離，m は音色空間の次数，S_A は分散・共分散行列を示す．同様の操作を他の音色カテゴリーに対して行うことにより，単音仮説がそれぞれの音色カテゴリーに属する確率を算出することができる．この確率値に基づいて，単音仮説生成時の初期確信度を与えた．

5.　単音仮説に基づくトップダウン処理

本章では，単音レベルと周波数成分レベルとの間の仮説コリレータの動作，すなわち単音に基づく周波数成分情報の付与について述べる．

5.1　知　識　源

単音情報に基づいて，出現する周波数成分を予測し，ある単音の下での周波数成分仮説に対する条件付確率の付与を行う．処理に際しては，単音に属する周波数成分を記憶蓄積した知識源（単音記憶：tone memories）を参照する．

単音記憶には，具体的な音の記憶として，次のようなデータを蓄積する．

$$T_k = \{a_{ij}\},$$
$$a_{ij} = (p_{ij},\ f_{ij}). \qquad (12)$$

すなわち，k 番目の単音記憶 T_k は，パワー値 p，周波数値 f を要素とする 2 次元ベクトル a_{ij} を要素とする行列である．但しパワー値は，その単音記憶中の周波数成分の最大パワー値によって正規化されており，また周波数値は，その単音記憶の基本周波数の時間平均値との比を表す．行列の各行（i）はその音に含まれる周波数成分に対応する．行の数は楽器種類によって 4〜20 程度である．また各列（j）は時間のサンプル点に対応する．本論文では 10 ms ごとに 1 サンプルとし，最大 80 のサンプル点をとった．単音記憶の概念図を図 4 に示す．このような単音記憶を，クラリネット，フルート，ピアノ，トランペット，バイオリンの 5 種類の自然楽器音について，音域別に蓄積した．

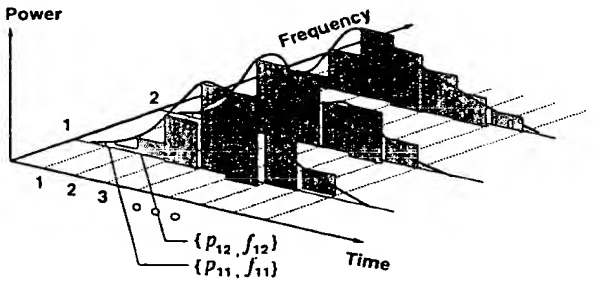


図 4　単 音 記 憶
Fig. 4　Tone memories.

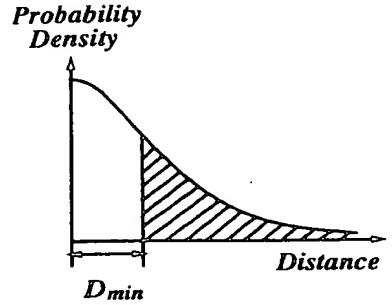
5.2　条件付確率の付与

単音仮説に基づいて周波数成分情報を付与する処理では，処理単位ごとに，あらかじめ蓄積した単音記憶を混合して，照合のための混合音（これを照合混合音という）を生成する．これは，式 (12) に示した単音記憶のうち，単音仮説中に存在する可能性のあるものを選び，それらを実際のパワー値と周波数値に換算し混合することによって行う．このようにして生成したすべての照合混合音について，その処理単位における周波数成分仮説に対する距離 D を求める．混合に際しては，単音記憶同士の立上り時刻のずれや，単音記憶に含まれる各周波数成分の位相差も考慮した．すなわち，立上り時刻や周波数成分の位相をそれぞれ変化させたときの距離 D の最小値 D_{min} をもって，その照合混合音と周波数成分仮説との間の距離とみなした．ここで距離 D は，

$$D = \sum_{i=1}^{F} \sum_{j=1}^{N} \left| p'_{ij} - p_{ij} \right| \cdot f_{ij} \qquad (13)$$

と定義した．ここで，F は照合する周波数成分の数（照合混合音と処理単位の周波数成分仮説において周波数が対応する周波数成分は一つと数える），N は照合混合音の時間軸方向のサンプル点の数，p'_{ij} は処理単位の周波数成分のパワー，p_{ij}，f_{ij} は照合混合音の周波数成分のパワーおよび周波数である．この距離に基づいて条件付確率値を定める．

距離から確率への変換においては，混合音の周波数成分の分布に関し，距離尺度上での正規分布を仮定する．すなわち，混合音の周波数成分は，距離尺度上で，照合混合音の周波数成分を中心とし一定の分散をもつ正規分布をなすと仮定する（図 5）．この場合，図 5 の網かけ部分の面積の 2 倍を，その照合混合音が与えられたときに，周波数成分が，照合した周波数成分仮説



混合音の周波数成分の分布に関し，距離尺度上で正規分布を仮定する．網かけ部分の面積の2倍を確率値とする．

図5　単音仮説の距離から確率値への変換
Fig. 5　Conversion of distance measure into probability.

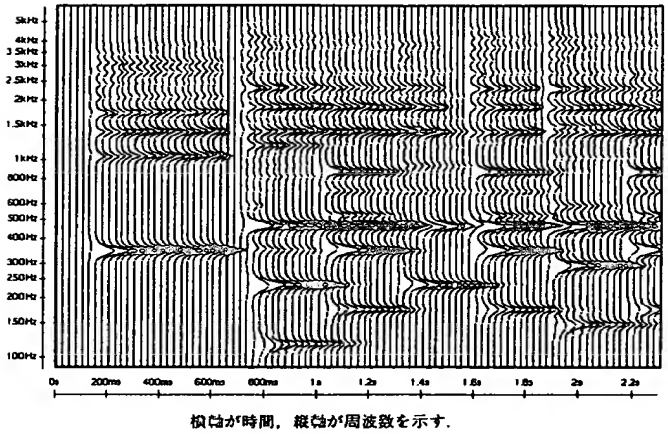


図6　周波数解析結果の例
Fig. 6　An example of spectrograms.

の状態となる確率（すなわち $P(C|N)$）とみなす．

6. システムの動作例

本章では，クラリネットとピアノによるアンサンブル演奏を入力した例を用いてシステムの動作を説明する．図6は，システムに入力された音響信号に対し周波数解析を行って得たスペクトログラムである．図7は，スペクトログラムに対して周波数成分を抽出した結果であり，周波数成分が線分として描かれている．図7中の縦の鎖線は，システムの前処理部で抽出した拍位置を示す．周波数成分のパワー値の変化と拍位置の情報を用いて，前処理部において処理単位が生成される．図7では，一例として入力開始から2番目の処理単位に属する周波数成分を太線で示している．処理単位が主処理部に入力されると，まず4.に述べた単音仮説クリエータが起動され，単音仮説生成される．単音仮説が生成されると，和音仮説クリエータと，5.に

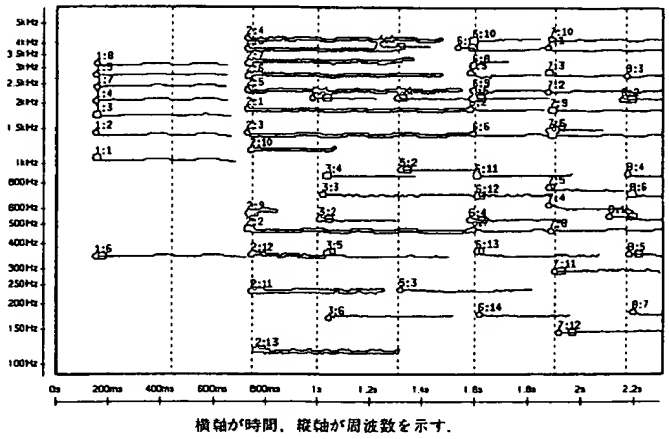


図7　抽出された周波数成分の例
Fig. 7　Frequency components extracted in the preprocessing block.

述べた単音レベルと周波数成分レベルとの間の仮説コリレータとが動作可能となるので，これらのモジュールが起動される．以下順次，図2に示したような仮説ネットワークが構成される．ネットワークが時間方向に一定の長さ（ノード数）に達すると，それより過去のノードが切り離され，仮説の状態が確定する．状態が確定した時点で，確信度最大の仮説がシステムから出力される．

7. 評価実験

本論文で提案する情報統合の機構に対し，単音レベルにおける情報統合の有効性を調べることを目的として，評価実験を行った．評価は，単音仮説生成モジュールだけを動作させた場合の単音認識結果（この場合，単音仮説のうちで初期確信度値が最も大きい仮説をシステムの認識結果とみなす）と，単音仮説生成モジュールと単音周波数成分情報付与モジュールの双方を動作させ情報統合を行った場合の単音認識結果とを比較することによって行う．

7.1 方　　法

評価用入力として，以下に述べる3種類の単音パターンを作り，これをサンプラで演奏した音響信号を作成した．ここでサンプラとは，任意の音響信号波形をそのままメモリに蓄積しておき，これを再生することによって発音する方式の音源装置である．本実験では，サンプラにクラリネット（cと表記），フルート（f），ピアノ（p），トランペット（t），およびバイオリン（v）の5種類の自然楽器音を蓄積した．単音パター

ンは，MIDI ノート番号 60 から 83 までの音域から一定数の音符（ここでは 2 音または 3 音）を選んで同時に発音する単音とし，これを時間方向に 50 個並べた．ここで同時とは，パターンを演奏する MIDI シーケンサ上で同じタイミングであることを意味する．各単音の継続時間は 750 ms とした．

単音パターンにおいては，各単音に由来する周波数成分の重なり方が単音認識の精度に大きく影響する．従って，ここでは単音パターンを次の三つの種類（クラス）に区別する．

（1） クラス 1 の単音パターン

同時に発音する単音の少なくとも 1 組が同一または整数倍の関係にある基本周波数をもつ（すなわちオクターブの関係にある）ような単音パターン

（2） クラス 2 の単音パターン

同時に発音する単音の少なくとも 1 組が 1.5 の整数倍の関係にある基本周波数をもつような単音パターンのうち，クラス 1 に属さないもの

（3） クラス 3 の単音パターン

クラス 1 にもクラス 2 にも属さない単音パターン

7.2 単音認識精度の指標

単音の認識精度の評価のための指標として，本論文では次に定義する正答指標 α，誤答指標 β，および認識率 R を用いる．

$$\alpha = \frac{a}{n}\ ,\ \beta = \frac{b}{n}\ ,\ R = \frac{1}{2}(\alpha - \beta) + \frac{1}{2} \qquad (14)$$

但し，n は入力（正解）に含まれる総音符数，a は出力に含まれる音符のうち音高と音色の両方が正しく認識された音符の数，b は出力に含まれる音符のうち，音高と音色のどちらかまたは両方が正しくない音符の数である．式 (14) における 1/2 の乗算と加算はスケール調整のためのものである．すなわち，システムが入力に含まれる音符と同数の音符を出力した場合，この正規化によって，すべてが誤っていれば $R = 0\%$，すべてが正しければ $R = 100\%$となる．

7.3 結　　果

図 8，図 9，および図 10 に，単音仮説生成の精度を示す．各図において，グラフは棒 2 本で 1 組となっており，左側の棒はボトムアップの単音仮説生成処理のみを動作させた場合の結果を，また右側の棒はボトムアップの単音仮説生成処理に加えトップダウンの単音周波数成分情報付与モジュールを動作させて，単音レベルでの情報統合を行った場合の結果をそれぞれ示している．グラフの横軸の表記においては，最初の数字が同時発音数，これに続くアルファベットが楽器音の種類を表す．例えば 2cp は，同時発音数 2 のパターンをクラリネットとピアノで演奏した場合の結果である．

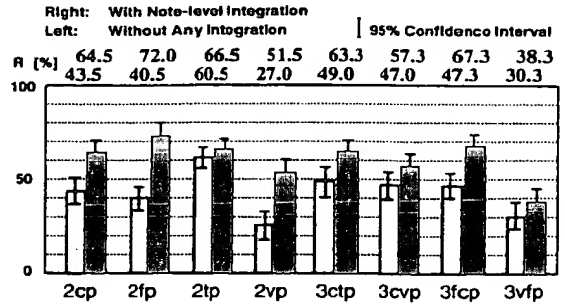


図 8　単音認識率の測定結果（クラス 1）
Fig. 8　Results of benchmark tests for note recognition (class 1).

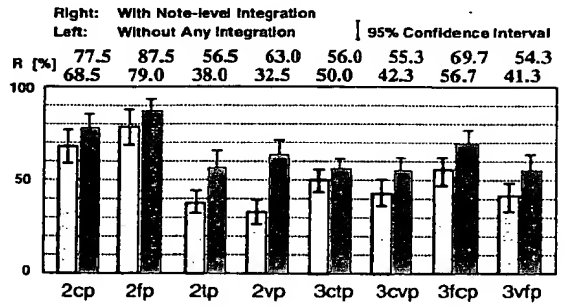


図 9　単音認識率の測定結果（クラス 2）
Fig. 9　Results of benchmark tests for note recognition (class 2).

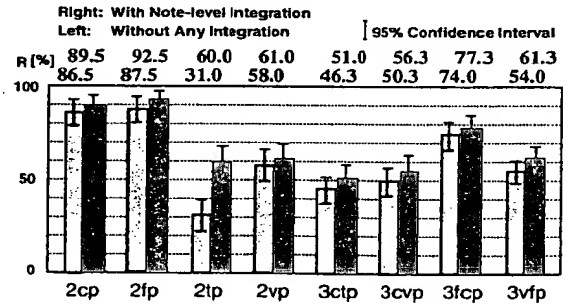


図 10　単音認識率の測定結果（クラス 3）
Fig. 10　Results of benchmark tests for note recognition (class 3).

図 11 に各クラスでの総音符数についての単音認識率を，また図 12 にクラス 1 の単音パターンに対する α と β をプロットしたグラフを示す．

認識率 R の値を見ると，各クラスとも，各グラフの左側に比較して右側の方が 3.0%から 31.5%（平均

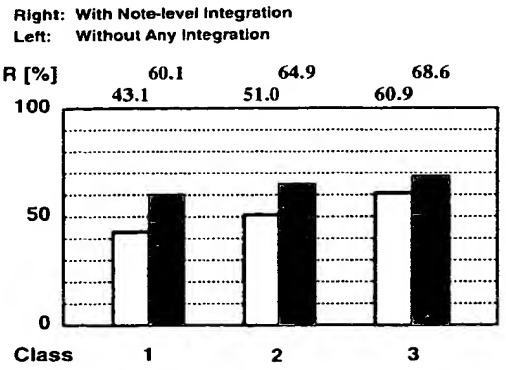


図 11 クラスごとの単音認識率
Fig. 11 Note recognition results for each class.

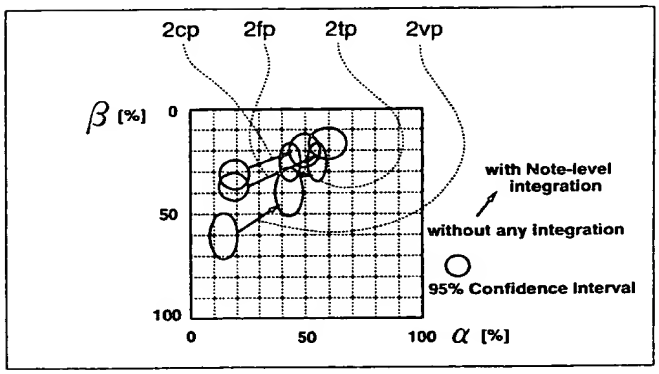


図 12 $\alpha-\beta$ 図（クラス 1）
Fig. 12 $\alpha-\beta$ plot (class 1).

12.9%）向上していることから，総じて単音レベルでの情報統合を行うことの効果が顕著に現れていると見ることができる．図 11 に示されるように，クラス別の認識率の向上の平均は，クラス 1 の場合 17.0%，クラス 2 の場合 13.9%，クラス 3 の場合 7.7%であり，周波数成分の重複する割合が大きくなるにつれて，単音記憶に基づく情報の統合の効果が大きくなっていることがわかる．

クラス 1（図 8）においては，少なくとも一つの単音の周波数成分は完全に他の単音の周波数成分と重複しているため，ボトムアップの単音仮説生成だけでは，ほとんど有効な処理結果を期待できない．実際，図 12 によれば，2cp，2fp，および 2vp において α 値が 20%以下にとどまっている．しかし，単音仮説生成部が基本周波数が整数倍の関係にある単音仮説を生成するため，単音構成周波数成分情報付与モジュールが有効に動作することができる．情報統合の結果，2cp，2fp，および 2vp では α 値，β 値ともに 20%程度，また 2tp でも 8%程度改善されている．

次にクラス 2（図 9）においては，同時に発音する単音において基本周波数成分は重複していないために，クラス 1 の場合に比べ単音認識率は一般に向上している．2fp，3fcp に対する結果に見るように，フルートの単音は比較的少数の周波数成分によって構成されるため，単音仮説生成において比較的高い認識率を得ている．

更にクラス 3（図 10）における実験結果を見ると，単音仮説生成の精度は，2cp，2fp，3fcp などのように，クラス 2 の場合よりも更に向上するものが多くなっている．

8. むすび

本論文では，音楽情景分析の処理モデル OPTIMA を提案し，ベイジアンネットワークによる情報統合の機構を示すと共に，単音の認識に最も関連の深い処理に絞って，情報統合の有効性を調べた．その結果，単音記憶の情報に基づくトップダウン処理を統合した場合，ボトムアップ処理のみの場合に比較して平均で 12.9%の単音認識率の向上が見られたことから，単音レベルにおける情報統合の有効性が示された．

しかし，本論文の実験で得られた単音認識率そのものは，いまだ実用的な値とは言えない．特に，トランペットのように基本周波数に対応する周波数成分のパワーが比較的低い楽器や，バイオリンのように周波数成分のパワーの著しい時間的変化や音色の揺らぎを有する楽器においては，クラリネットなど比較的定常な周波数成分を有する楽器に比べて単音認識精度が低い傾向が見られる．また，フルート演奏では息の音，クラリネット演奏ではピストンの操作音などが不可避的に混入するが，これらも単音認識精度を下げる一因となっている．また，ピアノのように，周波数成分が高調波関係であるという近似が，周波数が高くなるにつれて成り立たなくなる傾向にある楽器 [19] においては，単音形成処理における誤りも生じている．今後，これらの問題への対策が必要である．

本論文で述べた処理モジュールは，図 1 に示した構成要素のうちの一部分である．我々は，主処理部における和音レベルの情報統合についても有効性を示す実験結果を得ているので，これについて稿を改めて報告する予定である．

文　献

[1] A.S. Bregman, "Auditory Scene Analysis," MIT Press,

1990.
[2] M. Piszczalski and B.A. Galler, "Automatic music transcription," Computer Music Journal, vol.1, no.4, pp.24–31, 1977.
[3] 新原高水，今井正和，井口征士，"歌唱の自動採譜，" 計測論，vol.20, no.10, pp.940–945, 1984.
[4] B. Mont-Reynaud, "Problem-solving strategies in a music transcription system," Proc. of IJCAI85, pp.916–918, 1985.
[5] C. Roads, "Research in music and artificial intelligence," ACM Computing Surveys, vol.17, no.2, pp.163–190, 1985.
[6] 片寄晴弘，井口征士，"知的採譜システム，" 人工知能学会誌，vol.5, no.1, pp.59–66, 1990.
[7] 長束哲郎，才脇直樹，井口征士，"異種楽器を対象とした採譜システム，" 信学 '92 春大会，D-499, 1992.
[8] 植田　護，橋本周司，"ブラインドデコンポジション問題としての音源の分離と同定，" 情処研報 (93-MUS-3)，vol.93, no.93, 1993.
[9] G.J. Brown and M. Cooke, "Perceptual grouping of musical sounds: A computational model," Journal of New Music Research, vol.23, pp.107–132, 1994.
[10] 柏野邦夫，"計算機による聴覚の情景分析 — はじめの一歩，" 日本音響学会誌，vol.50, no.12, pp.1023–1028, 1994.
[11] S. Handel, "Listening," MIT Press, 1989.
[12] W.M. Hartmann, "Pitch Perception and the Segregation and Integration of Auditory Entities," in G.M. Edelman, et al. (eds.), "Auditory Function, Neurobiological Bases of Hearing," pp.623–645, John Wiley & Sons, 1988.
[13] T. Nakatani, H.G. Okuno, and T. Kawabata, "Auditory stream segregation in auditory scene analysis with a multi-agent system," Proc. of 12th National Conf. on Artificial Intelligence, pp.100–107, 1994.
[14] 中谷智広，奥乃　博，川端　豪，"音環境理解のためのマルチエージェントによる調波構造ストリームの分離，" 人工知能学会誌，vol.10, no.2, pp.232–241, 1995.
[15] V. Lesser, S.H. Nawab, I. Gallastegi, and F. Klassner, "IPUS: An architecture for integrated signal processing and signal interpretation in complex environments," Proc. of 11th National Conf. on Artificial Intelligence, pp.249–255, 1993.
[16] J. Pearl, "Fusion, propagation, and structuring in belief networks," Artificial Intelligence, vol.29, no.3, pp.241–288, 1986.
[17] 柏野邦夫，田中英彦，"2 つの周波数成分の分離知覚に関する工学的モデル — 複数の要因の評価と統合，" 信学論 (A)，vol.J77-A, no.5, pp.731–740, 1994.
[18] 山口公典，安藤繁雄，"短時間スペクトル分析法の自然楽器音への適用，" 日本音響学会誌，vol.33, no.6, pp.291–300, 1977.
[19] 安藤由典，"新版楽器の音響学，" 音楽之友社，1996.

（平成 7 年 10 月 20 日受付，8 年 4 月 9 日再受付）

柏野　邦夫　（正員）

平 2 東大・工・電子卒．平 7 同大大学院電気工学専攻博士課程了．工博．同年 NTT に入社，基礎研究所情報科学研究部勤務，現在に至る．聴覚的情景分析の研究に従事．音響的情報を対象とする信号処理および知識処理に興味をもつ．平 6 情報処理学会奨励賞受賞．情報処理学会，人工知能学会，日本音響学会，IEEE 各会員．

中臺　一博　（正員）

平 5 東大・工・電気卒．平 7 同大大学院情報工学専攻修士課程了．同年 NTT に入社，ソフトウエア本部勤務，現在に至る．在学中，聴覚的情景分析の研究に従事．情報処理学会，人工知能学会，日本音響学会各会員．

木下　智義　（学生員）

平 7 東大・工・電子情報卒．現在同大大学院情報工学専攻修士課程在学中．聴覚的情景分析の研究に従事．情報処理学会会員．

田中　英彦　（正員）

昭 40 東大・工・電子卒．昭 45 同大大学院博士課程了．工博．同年東大・工・講師，昭 46 同大助教授，昭 62 同大教授，現在に至る．この間昭 53〜54 ニューヨーク市立大客員教授．計算機アーキテクチャ，並列推論マシン，帰納推論，オブジェクト指向計算システム，分散処理，CAD 等の研究に従事．著書「非ノイマンコンピュータ」，「情報通信システム」，共著書「計算機アーキテクチャ」，「VLSI コンピュータ I, II」，「ソフトウエア指向アーキテクチャ」．情報処理学会，人工知能学会，日本ソフトウエア科学会，IEEE，ACM 各会員．

音楽情報科学 19－11
(1997. 2. 21)

適応型混合テンプレートを用いた音源同定
－複数楽器演奏への適用－

柏野 邦夫　　村瀬 洋

NTT基礎研究所
〒243-01 厚木市森の里若宮 3-1

kunio@ca-sun1.brL.ntt.co.jp, murase@apollo3.brL.ntt.co.jp

あらまし 同時に複数の認識対象が混在する音の認識では，音源同定処理が必要である．本稿では，音楽の生演奏など，実環境における音の多様性や変動にも対処できる音源分離同定を行うことを目的として，適応型テンプレートを用いた音源同定処理を提案する．さらに，この処理を応用して，同時に複数の音を認識対象とするシステムの代表例であるアンサンブル演奏の認識システムを構築する．構築したシステムに対し，自然楽器音の単音によるベンチマークテスト，およびアンサンブルの生演奏を用いた音楽認識テストを行った結果，単純なマッチトフィルタによる音源同定処理に比べ，提案手法が有効であることが確かめられた．

キーワード 聴覚的情景分析，音源同定，音源分離，音楽情景分析，自動採譜，マッチトフィルタ

Sound Source Identification Using Adaptive Template Mixtures
– Formulation and Application to Music Stream Segregation –

Kunio Kashino and Hiroshi Murase

NTT Basic Research Laboratories
3-1 Morinosato-Wakamiya, Atsugi-shi,
243-01, Kanagawa, Japan.

Abstract Sound source identification is an essential problem in auditory scene analysis when multiple acoustical objects are simultaneously present in the scene. However, little work has been done on sound source identificaiton for a multiple-source environment. Here we propose a novel method for sound source identification. The key idea is adaptation of templates, which has permitted to cope with variation of sounds. As an example application of the proposed method, we have built a music recognition system that recognizes instrument names and pitches of the notes included in ensemble music performances. Experimental results show that the proposed adaptive mechanisms significantly improve the accuracy of sound source identification in comparison to a conventional matched-filter-based method.

key words auditory scene analysis, sound source identification, sound source separation, music scene analysis, automatic music transcription, matched filter

1 まえがき

音の認識の研究では，従来，認識の対象とする音の種類をあらかじめただ一つに限定するものがほとんどであった．例えば音声認識システムは，その名の通り人間の音声だけを認識の対象とする．もちろん，音声認識システムの入力として，色々な雑音が混在していることを考慮することも多いが，その場合も，認識の対象となるのは音声だけである．

これに対しわれわれは，複数種類の認識対象が混在する場合の音の認識に取り組んでいる．この問題は，シグナル（認識対象の音）とノイズ（認識対象ではない音）が一義的に決まっているのではなく，同時に複数の音がシグナルとなり得るのが特徴である．このような問題は，音によって周囲の状況を理解しようとする場合はもちろん，音声認識のように一見認識対象が決まっている場合であっても，実環境での自然なヒューマンインタフェースの手段として利用できるような完成度の高いシステムの構築を目指すのであれば，避けて通ることのできない問題である．

さて，複数種類の認識対象が混在する音の認識では，入力の音響信号から個々の音に相当する部分を分けて取り出すことと，個々の音が何の音であるかを判定することの二つの課題がある．本稿では，前者を音源分離（sound source separation），後者を音源同定（sound source identification）と呼ぶ．従来の auditory scene analysis（聴覚的情景分析）の工学的研究では，主として音源分離だけが議論されてきた．一方，室内などの音響事象の認識 [1]，話者認識，音声区間の切り出しなど，単一で存在する音源の識別の研究も行われている．しかし，同時に複数種類の音が存在する状況下でそれぞれの音源を同定する問題は，これまでほとんど検討されていない．

そこで本稿では，音源同定の問題を扱う．音源同定に関する研究としては，柏野らによる OPTIMA の研究がある [2, 3]．OPTIMA は，情報統合を鍵技術とする音楽認識の処理モデルである．これまでに，2～3パートの編成のモノラルのアンサンブル演奏を入力とし，種々の情報を統合してパートごとの音符情報などを出力する実験システムが実装されている．しかしながら，その評価実験は主にサンプラの音を用いて行われていた．これは，生楽器音は多様で変動が大きいために，精度良く処理することが難しかったからである．

本稿では，多様で変動の大きい対象を扱うための鍵技術として「適応」の考え方を導入する．以下 **2.** では，音源同定のためのテンプレートを入力に合わせて変化させるという適応型混合テンプレートのアイデアを提案し，問題の定式化を行う．**3.** では，計算を実行するための具体的なシステムの構成を議論する．**4.** では，構築したシステムに対し簡単な評価実験を行って，**2.** で提案する処理の有効性を検討する．**5.** をむすびとする．

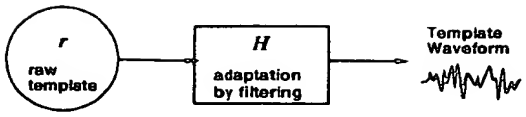


図 1: 音源波形のテンプレートフィルタリングモデル

2 適応型テンプレート

2.1 テンプレートフィルタリング

各音源波形の和で与えられる波形を各音源波形に分離する問題を解くための制約として，各音源波形のモデル $y_n(k)$ を与えることを考える．ここで n は各音源に対応する添字，k はサンプル時刻を表す．すると，われわれの問題は，

$$J = E\left[\left\{z(k) - \sum_{n=0}^{N-1} y_n(k)\right\}^2\right] , \qquad (1)$$

の最小化として定式化することができる．ここで $z(k)$ は入力信号波形，N は音源の数，E は時間平均を表す．なお N はあらかじめ与えられてはいない．音源波形 $y_n(k)$ のモデルとして，図 1 に示すような「テンプレートフィルタリングモデル」を考える．これは，ある一群の音源波形を，原テンプレートと線形フィルタによる変形とでモデル化するものである．線形フィルタとして FIR 型を用いることにすれば，

$$y_n(k) = \sum_{m=0}^{M-1} h_n(m)\, r_n(k-m) , \qquad (2)$$

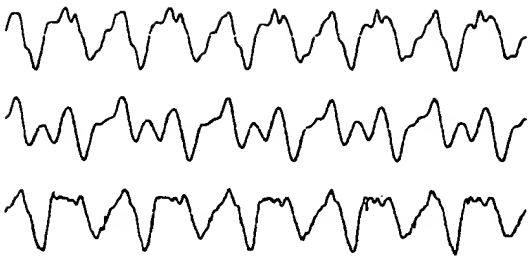
と書ける．ここで，h は FIR フィルタのインパルス応答，r は原テンプレート波形，M はフィルタの次数である．

一般に音源波形は多様であり変動するので，h や r として固定の値を用いることはできない．音源波形の多様性の例を図 2 に示す．もし位相を捨てて例えばパワースペクトル表現を用いることにしても，その表現の空間上で音が変動するという事情は基本的に同じである．したがって，音源の変動に対処する何らかの仕組みが必要である．ここでは，フィルタの係数 $h_n(m)$ を変えることを考える．式 (1) を，式 (2) を用いて書き直すと

$$J = E\left[\left\{z(k) - \sum_{n=0}^{N-1}\sum_{m=0}^{M-1} h_n(m)\, r_n(k-m)\right\}^2\right] , \qquad (3)$$

となる．

この J が $h_n(m)$ に関して最小となるための必要条件は，全ての n と m に関して，偏微分 $\partial J/\partial h_n(m)$



上段はベーゼンドルファ，中段はヤマハのピアノである．どちらも同じ高さ (F4), 同じ時間部分 (立上りから 100ms ～130ms) であり，ほぼ同じ強さで弾いたものであるが，波形は異なっている．しかし，適切な FIR フィルタを通すことによって，中段の波形を上段の波形にあわせて変形させることができる．下段は，40 次の FIR フィルタによって変形させた中段の波形．上段の波形にかなり近づいている．サンプリング周波数は 48kHz.

図 2: ピアノ波形の多様性とその吸収

が 0 となることである．この条件を用いると，$N \times M$ 個の連立一次方程式

$$\sum_{n=0}^{N-1} \sum_{m=0}^{M-1} E\left[r_i(k-l)\, r_j(k-m)\right]\, h_n(m) = E\left[r_i(k-m)\, z(k)\right] \quad . \tag{4}$$

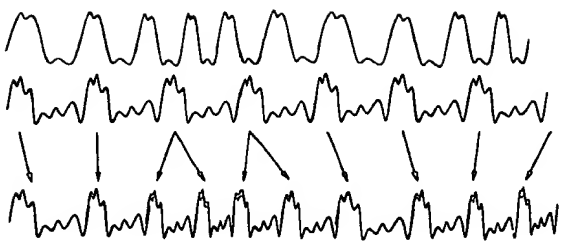
を導くことができる．未知数の個数と方程式の個数が等しいので，この連立方程式は，係数行列の逆行列を求めることによって解くことができる．式 (4) の係数行列は，$i \neq j$ のとき $r_i(k)$ が $r_j(k)$ の定数倍とならないように定めておけば，正則となる．

2.2 位相トラッキング

前節の処理が有効となるためには，原テンプレート r の基本周波数および位相が，z に含まれている音源の基本周波数および位相と一致していなくてはならない．なぜなら，フィルタ H は，信号の周波数を変えることはできないからである．このため，r の位相を，z 中の対応する音源の位相に時々刻々合わせ込むメカニズムが必要である．

もし，入力信号が，複数の音源からの音が混在したものではなく，一つの音源からの音であれば，既に提案されている適応ピッチトラッキングの手法を用いることができるであろう [5]. しかし，そのような信号処理の手法は，そのままでは混合音に対して適用することはできない．そこでわれわれは，混合音に対して適用できる位相トラッキングの手法を考案した．これは，次の 6 ステップから成る．

(1) 入力信号 z に対して周波数解析を行い，基本周波数成分を全て抽出する．z は複数の音源からの音の混合物かも知れないから，複数の基本周波数成分があるかも知れないことを考慮する．ただし，複数の音源の基本周波数が整数倍の関係にあることは考えない．



上段：入力波形 z；中段：位相トラッキング前の原テンプレート；下段：位相トラッキング後の原テンプレート．下段の波形が式 (4) における原テンプレート $r_i(k)$ として用いられる．なお本図は説明図であり処理結果を示したものではない．

図 3: 位相トラッキングの説明図

(2) 抽出された各基本周波数について，対応する音源であるかも知れない原テンプレート r_i を選び出す．

(3) 狭帯域のバンドパスフィルタを r_i に適用する．バンドパスフィルタの中心周波数は，それぞれの r_i の平均的な基本周波数とする．バンドパスフィルタの出力は，正弦波に近い波形となるので，その位相をバッファに保持する．位相の時系列を $p_{r,i}(k)$ とおく（k は時刻）．

(4) r_i に対して適用したのと同じバンドパスフィルタを入力信号 z に対して適用し，(3) と同様に位相 $p_{z,i}(k)$ を保持する．

(5) 入力波形とテンプレート波形の時々刻々の時間差 $\delta k_{r,i}(k)$ を求める．位相差 $\delta p_{r,i}(k)$ は

$$\delta p_{r,i}(k) = p_{z,i}(k) - p_{r,i}(k), \tag{5}$$

で与えられるから，時間差 $\delta k_{r,i}(k)$ は

$$\delta k_{r,i}(k) = \frac{f_s}{2\pi f_{c,i}}\, \delta p_{r,i}(k), \tag{6}$$

によって計算できる．ここで f_s はサンプリング周波数，$f_{c,i}$ は適用されたバンドパスフィルタの中心周波数である．

(6) 時刻 k での位相トラッキング後の波高値 $r_i(k)$ は，求められた時間差を用いて

$$r_i(k) = r_i(\, k - \delta k_{r,i}(k)\,) \tag{7}$$

によって求めることができる．

図 3 は，上記のアルゴリズムが動作する様子を表した説明図である．

以上述べたように，本手法は，音源の変動を基本周波数のゆらぎと，基本周波数に対する高調波の相対位相や振幅の変動による波形の歪みとに分けて考え，前者を位相トラッキングによって，また後者をテンプレートフィルタリングによって吸収するものである．

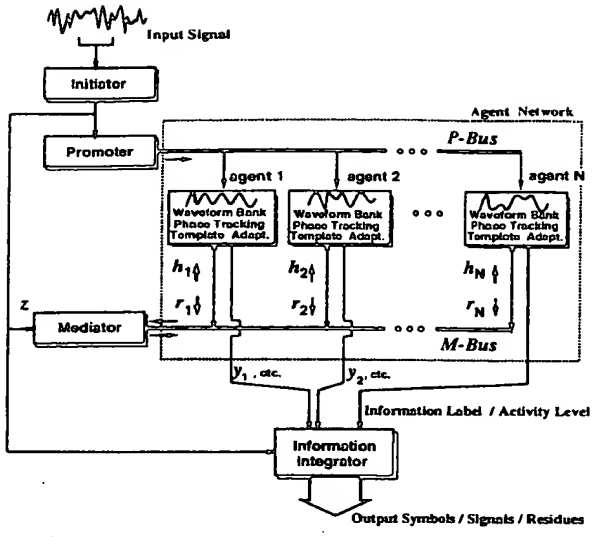


図 4: 提案する Ipanema アーキテクチャ

3 マルチエージェントアーキテクチャによる実装

本節では，前節で導入した計算を行うためのシステムの構成について議論する．

3.1 概要

同時に複数の認識対象の音が存在し得るとき，ある音をシグナルと捉え他の音をノイズとみなすような処理モジュールを複数準備しておき，それらを並列に動作させることによって個々の音の認識を図るのは，きわめて自然な発想であろう．それぞれの処理モジュールは，各々が担当する音だけを検出するという比較的単純な機能をもち，また処理モジュールは全く独立にではなく，相互に影響を及ぼしながら動作する．これはマルチエージェントアーキテクチャの考え方に他ならない．

図 4 に，提案するシステムの処理モデル（アーキテクチャ）を示す．このシステムは，複数種類の音が混在した音響信号を入力とする．本稿の範囲では，入力信号は音楽演奏である．出力としては，楽譜に類似した形式の記号表現，各音源ごとの音響信号，および解釈の残差の音響信号を生成する．

図 4 のアーキテクチャは，処理のきっかけを与えるイニシエータ（initiator），エージェントの処理を先導するプロモータ（promoter），音源分離・同定処理の主体となるエージェントネットワーク（agent network），およびエージェントの調停を行うメディエータ（mediator of agents）から成っている．そこで，図 4 のアーキテクチャを「Ipanema」と呼ぶ．また，上記の要素の他に，後処理モジュールとして情報インテグレータ（information integrator）が備わっている．

3.2 処理モジュール

3.2.1 イニシエータ

イニシエータは，入力信号を受け取り，これをフレーム（ある時間範囲）ごとに切り出した波形を出力する．イニシエータの出力は，後続の処理のきっかけとなる．フレーム長は一定ではなく，音の立上りを検出するごとに新たなフレームを生成する．

3.2.2 プロモータ

プロモータは，1 フレームの波形を受け取り，周波数解析を行って，フレーム中に含まれている基本周波数成分を抽出する．フレーム中に複数の音が混在している場合には，基本周波数成分も複数存在する．プロモータは，抽出した基本周波数と，パワー包絡など音のおおまかな特徴量とを，プロモーションバス（P-Bus）に書き出す．この情報を P-Bus 情報と呼ぶ．P-Bus は，プロモータによって書き込まれ，次に述べるエージェントたちによって読み出される共通のデータ領域である．P-Bus 情報は，各エージェントが活動するかどうかを決めるために用いられる．

3.2.3 エージェント

Ipanema アーキテクチャでは，エージェントネットワーク中の一つのエージェントは，個々の音源種類（例えばフルート）に対応している．エージェントは，原テンプレートの集合であるテンプレートバンクを持っている．テンプレートバンク中には，例えば半音ずつピッチの異なる単音の波形が原テンプレートとして蓄積されている．

各エージェントは，随時 P-Bus を観察しており，プロモータによって書き出された P-Bus 情報を読み出す．P-Bus 情報中の基本周波数およびパワー包絡など単音のおおまかな特徴量が，自分の担当する音源種類の基本周波数および特徴量と矛盾しなければ，エージェントは担当音源が入力に含まれている可能性があるものと判断して活動状態となる．すなわち，テンプレートバンクから，基本周波数や特徴量が現在の入力と最も整合する波形を選び出し，前節で述べた位相トラッキング処理を行って原テンプレート r_i を生成する．もし，P-Bus 情報中の基本周波数や特徴量が自分の担当する音源種類と矛盾すれば（たとえば発音不可能な音域であるなど），そのエージェントは何もせず，次の P-Bus 情報が準備されるまで休眠する．

活動状態のエージェントから生成された原テンプレート r_i は，メディエーションバス（M-Bus）と呼ばれる共通のデータ領域に書き出される．M-Bus は，エージェントや次項に述べるメディエータによって読み書きされる共通のデータ領域である．エージェントが書き出した r_i は，メディエータによって処理される．メディエータからは，各エージェントに対応するフィルタ係数が返って来るので，各エージェントは，そのフィルタ係数を M-Bus から読み込んで，r_i に対

してフィルタ演算を行う．これによって **2.** に述べたテンプレートフィルタリングが実現される．

エージェントからの最終的な出力は，テンプレートフィルタリングの結果としての波形 y_i，活動レベルとしての平均パワー $E[y_i^2]$ および記号表現のラベル（例えば「ピアノの C4」）である．

M-Bus に関しては，現在の実装では，エージェントはメディエータに対してのみ情報を渡し，またメディエータからのみ情報を受け取る．しかし将来的には，M-Bus を介して任意のエージェントが任意のエージェントに対して情報を受け渡すような処理形態も考えられる．この場合 M-Bus は，黒板モデルにおける黒板の役割を果たしていると見ることもできる．

3.2.4 メディエータ

メディエータは，各エージェントの出力を調整する役割を負う．本稿においては，各エージェントの提案する原テンプレートに対するフィルタ係数を返すことによって出力の調整が行われる．すなわちメディエータは，イニシエータから入力波形のフレーム z が切り出されてから一定時間待ち，その時間内に M-Bus に書き込まれた原テンプレート r_i を読み込む．これらに基づいて，連立方程式 (4) を解けば，各エージェントに対するフィルタ係数 h_i が得られるので，これを M-Bus に書き込んでエージェントに返す．

3.2.5 情報インテグレータ

情報インテグレータは，システムの出力の一部を修正する後処理モジュールである．情報インテグレータは，各エージェントから，波形 y_i，活動レベル $E[y_i^2]$ および記号表現のラベルを受け取る．現在の実装は簡略化されていて，情報インテグレータはただ活動レベルに基づいて実際に発音している音源を判定するのみである．したがって現状では情報のインテグレーションという名前と実体とは整合していない．

しかし，近い将来このモジュールは拡張される予定である．これまでに述べた Ipanema の枠組では，フレーム単位に処理が行われるために散発的に発音判定の誤りが生じることがある．これに対しては，音どうしの継時的あるいは同時的関係に着目して出力判定を修正することがきわめて効果的である．この修正とは，結局，統計的情報など対象に関する複数の知識を統合して最も確からしい判断を行うことである．このようなわけで，本モジュールは情報インテグレータと名付けられている．情報のインテグレーションの具体的な手法としては，Bayesian network を応用した OPTIMA [2, 3] が一つのベースになり得ると考えている．

4 評価実験

提案法による認識精度を評価するため，単音認識のベンチマークテストと音楽認識テストを行った．

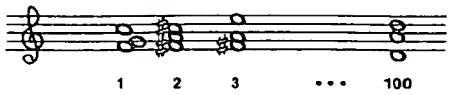


図 5: ベンチマークテストに用いる単音パターンの例

表 1: 実験に用いた楽器

	テンプレート	テストパターン
ピアノ	ベーゼンドルファ 225	ヤマハ C2
バイオリン	ハンニバル ファグノラ	作者不詳（1720 年 クレモナ製）
フルート	ブランネンクーバー	アルタス（頭部）＋ムラマツ

4.1 ベンチマークテスト

ここで用いるベンチマークテストは，文献 [2] で行ったものと同様のものである．用いたテストパターンは，図 5 に示すような 3 つの単音が同時に鳴るパターンである．パターンはクラス 2 とした．クラス 2 とは，同時に発音する単音の少なくとも一組が 1.5 の整数倍の関係にある基本周波数を持つような単音パターンのことである [2].

パターンの作成においては，あらかじめフルート，ピアノ，およびバイオリンの自然楽器の単音を半音ごとにスタジオで収録した（16bit, 48kHz）．この波形を計算機上に蓄積しておき，これをクラス 2 および MIDI ノート番号 60 ～74 という制約の中でランダムに選択して加算することによってパターンを作成した．

認識率 R の定義は，文献 [2] などと同様に

$$R = 100 \cdot \left(\frac{right - wrong}{total} \cdot \frac{1}{2} + \frac{1}{2} \right) \quad [\%], \qquad (8)$$

とした．ただし $right$ は出力に含まれる音符のうち音高と音色の両方が正しく認識された音符の数，$wrong$ は出力に含まれる音符のうち，音高と音色のどちらかまたは両方が正しくない音符の数，$total$ は入力 (正解) に含まれる総音符数である．予備実験の結果から，テンプレートフィルタリング On の条件においては，FIR フィルタの次数を 40 とした．なおテンプレートフィルタリング Off とは，FIR フィルタの次数を 1 としたという意味である．

本実験では，原テンプレートとしてテストパターンの生成に利用するのと同一の波形を用いたり，同一個体の楽器を用いるたりすると，波形の一致度が高いために評価実験としては適切でない．そこで，テンプレートの波形とテストパターンの波形は，互いに異なる個体から収録したものを用いた．これを表 1 に示す．

表 2: ベンチマークテストの結果

		テンプレートフィルタリング	
		On	Off
位相トラッキング	On	77.3 % ± 4.1 %	64.7 % ± 4.9 %
	Off	61.0 % ± 4.5 %	57.8 % ± 4.8 %

± は 95 % 信頼区間を示す.

表 3: 音楽認識テストの結果

		テンプレートフィルタリング	
		On	Off
位相トラッキング	On	66.3 %	61.0 %
	Off	52.7 %	52.3 %

実験結果を表 2 に示す. この表では, 右下の欄の条件 (テンプレートフィルタリング Off, 位相トラッキング Off) が, 単純なマッチトフィルタによる音源同定に相当している. したがって, マッチトフィルタに比較して, **2.** で提案した適応型テンプレートを用いる処理の有効性が明確に示されていると見ることができる.

4.2 音楽認識テスト

ベンチマークテストに加え, 音楽の生演奏を対象とした音楽認識テストを行った. ここでは, 表 1 とはまた別の楽器個体のバイオリン, フルート, ピアノを用いて演奏したテスト曲「蛍の光」を対象として, 音源同定処理についての認識率 R を調べた. テスト曲の楽譜は [3] のものを用いた. 表 3 にその結果を示す. 表中の値は音源同定処理だけに関する認識率である. 結果の定性的傾向はベンチマークテストと同様であり, 提案手法の有効性が示されている.

5 むすび

本稿では, 音楽の生演奏など, 実環境における音の多様性や変動にも対処できる音源同定を行うことを目的として, 適応型テンプレートを用いた音源同定処理を提案した. さらに, この処理を応用して, 同時に複数の音を認識対象とするシステムの代表例であるアンサンブル演奏の認識システムを構築した. このシステムは, マルチエージェントアーキテクチャに基づくことにより, モジュラリティとスケーラビリティを持たせた形で簡明に実装することができた. 自然楽器音の単音によるベンチマークテスト, およびアンサンブルの生演奏を用いた音楽認識テストの結果, 単純なマッチトフィルタによる音源同定処理に比べ, 提案手法が有効であることが確かめられた.

従来, マルチエージェントベースの音響処理システム [1, 4] では, エージェント間の通信は明確に定式化されないことが多かった. これに対し本稿のシステムでは, エージェント間の通信を二乗平均誤差の最小化という規範で定量化している点が特徴である. また, これまでに, 複数種類の楽器のアンサンブル演奏を扱うことのできる音楽認識システムも提案されているが [2], 生演奏を扱うことは容易ではなかった. これに対し本稿は, 生のアンサンブル演奏に対する認識の可能性を示したものと位置付けることができる.

しかしながら, 認識精度自体はまだ実用的なものではない. 今後の課題として, まず情報インテグレータをフルに実装した形での評価が必要である. 次に, テンプレートフィルタのパラメトライズの問題がある. すなわち FIR フィルタだけではなく, 各音源のもつバリエーションをうまく表現するようなフィルタを構成することは, 高精度化を図る上で重要な課題である. このためには, 対象の多様性や変動をミクロにモデル化することも必要となろう. 一方でわれわれは, 適応型混合テンプレートの枠組を, 音楽以外の例題に対して適用することも試みたいと考えている.

謝 辞

議論して頂いた, 弊社基礎研究所の奥乃 博 主幹研究員, 川端 豪 主幹研究員, 中谷 智広 研究主任に感謝する. また日頃サポートを頂く同研究所情報科学部の石井健一郎部長に感謝する.

参考文献

[1] Lesser V., Nawab S. H., Gallastegi I. and Klassner F. IPUS: An Architecture for Integrated Signal Processing and Signal Interpretation in Complex Environments. In *Proceedings of the 11th National Conference on Artificial Intelligence*, 249–255, 1993.

[2] 柏野 邦夫, 中臺 一博, 木下 智義, 田中 英彦: 音楽情景分析の処理モデル OPTIMA における単音の認識. 信学論 D-II, **J79-DII**, 11, 1751–1761, 1996.

[3] 柏野 邦夫, 木下 智義, 中臺 一博, 田中 英彦: 音楽情景分析の処理モデル OPTIMA における和音の認識. 信学論 D-II, **J79-DII**, 11, 1762–1770, 1996.

[4] 中谷 智広, 後藤 真孝, 川端 豪, 奥乃 博: 残差駆動型アーキテクチャの提案と音響ストリーム分離への応用. 知能誌, 12, 1, 111–119, 1997.

[5] Nehorai A. and Porat B.: Adaptive Comb Filtering for Harmonic Signal Enhancement. *IEEE Trans. on ASSP*, 34, 5, 1124–1138, 1986.